



Analysis of Factors that Influence Video Engagement on the TikTok Platform Using the Multiple Linear Regression Algorithm

Analisis Faktor-Faktor yang Mempengaruhi *Engagement* Video di Platform *TikTok* Menggunakan *Multiple Linear Regression*

Nur Sapina^{1*}, Annisa Nanda², Muhammad Amirul Arifin³,
Rahmaddeni⁴, Lusiana Efrizoni⁵

^{1,2,3,4,5}Program Studi Teknik Informatika, Universitas Sains dan Teknologi Indonesia, Indonesia

E-Mail: ¹2310031802147@sar.ac.id, ²2417052802081@usti.ac.id, ³2417052802086@usti.ac.id,
⁴rahmaddeni@usti.ac.id, ⁵lusiana@stmik-amik-riau.ac.id

Received Feb 27th 2025; Revised Apr 21th 2025; Accepted May 26th 2025; Available Online Jun 24th 2025, Published Jun 24th 2025
Corresponding Author: Nur Sapina

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

TikTok has grown into one of the most widely recognized digital engagement platforms worldwide, with over one billion active users. However, some videos on TikTok achieve high engagement rates despite using similar content approaches. This research is intended to explore the elements that influence video engagement on TikTok by applying the Multiple Linear Regression algorithm. The variables analyzed include video duration, number of views, comments, likes, shares, and downloads. After data preprocessing, feature selection, and regression assumption testing, it was found that video_like_count, video_share_count, and video_download_count have the most significant influence on view count. The model evaluation results proved that the regression model showed excellent predictive performance, with an R^2 Squared value of 0.978, RMSE of 0.0742, and MSE of 0.0055. This research provides practical insights for content creators and content marketers in designing more optimized content production. This prediction model can also be utilized to estimate the engagement potential of a video before it is published.

Keywords: Content Prediction, Engagement, Multiple Linear Regression, Social Media, TikTok

Abstrak

TikTok telah berkembang menjadi salah satu platform interaksi digital terkenal secara luas di seluruh dunia, yang memiliki lebih dari satu miliar orang pengguna aktif. Namun, sebagian video di TikTok memperoleh tingkat engagement yang tinggi meskipun menggunakan pendekatan konten yang serupa. Riset ini dimaksudkan untuk menelusuri unsur-unsur yang memberikan pengaruh terhadap engagement video di TikTok dengan menerapkan algoritma Regresi Linear Berganda. Variabel yang dianalisis meliputi durasi video, jumlah tayangan, komentar, like, share, dan download. Setelah melalui tahap preprocessing data, seleksi fitur, dan pengujian asumsi regresi, ditemukan bahwa video_like_count, video_share_count, dan video_download_count memiliki pengaruh paling signifikan terhadap jumlah tayangan. Hasil evaluasi model membuktikan bahwa model regresi menunjukkan kinerja prediktif yang sangat baik, dengan nilai R^2 Squared sebesar 0,978, RMSE sebesar 0,0742, dan MSE sebesar 0,0055. Riset ini memberikan gambaran praktis kepada konten kreator dan konten marketing dalam merancang produksi konten yang lebih optimal. Model prediksi ini juga dapat dimanfaatkan untuk memperkirakan potensi engagement suatu video sebelum dipublikasikan.

Kata Kunci: Engagement, Media Sosial, Prediksi Konten, Regresi Linear Berganda, TikTok

1. PENDAHULUAN

TikTok adalah satu diantaranya layanan jejaring sosial yang berfokus pada video pendek dan tengah mengalami pertumbuhan signifikan secara global, termasuk di Indonesia. Berdasarkan data Januari 2023, jumlah pengguna *TikTok* secara global mencapai sekitar 1,05 miliar, dengan Amerika Serikat sebagai negara dengan jumlah pengguna terbanyak, yaitu 113,25 juta, diikuti oleh Indonesia dengan 109,90 juta pengguna [1]. *TikTok* memberikan kesempatan bagi user menciptakan rekaman visual pendek menggunakan efek inovatif serta berbagai efek visual yang atraktif dengan demikian mengundang ketertarikan penonton [2]. Hal ini



membuat TikTok menjadi platform yang menarik bagi individu, merek, serta pemasar digital dalam menjangkau audiens secara luas.

Satu diantara aspek penting yang berperan besar dalam kesuksesan konten di TikTok adalah engagement, yang mencerminkan tingkat interaksi audiens terhadap sebuah video. *Engagement* media sosial adalah menentukan total keterlibatan yang dipunyai user melalui video yang diunggah di media sosial. Setiap suka, bagikan, komentar, *retwit* atau sebutan adalah bentuk keterlibatan yang menunjukkan bahwa pengguna tertarik dengan konten tersebut. Memiliki *engagement* media sosial yang lebih baik berarti pelanggan memiliki hubungan yang lebih kuat dengan merek [3]. Tingginya engagement menandakan bahwa sebuah video berhasil menarik perhatian dan membangun koneksi dengan audiens. Dalam konteks pemasaran digital, engagement yang tinggi dapat meningkatkan eksposur merek, memperkuat hubungan dengan pelanggan, serta mendorong konversi bisnis

Namun, tidak semua video di *TikTok* mendapatkan tingkat *engagement* yang sama, meskipun menggunakan pendekatan atau tema yang serupa. Berbagai faktor, seperti waktu pengunggahan, panjang video, kualitas konten, dan interaksi dengan *audiens*, dapat mempengaruhi seberapa banyak perhatian yang diterima oleh sebuah video. maka dari itu, esensial untuk meneliti aspek-aspek yang berdampak pada *engagement* video di *TikTok* untuk memahami dinamika dan karakteristik yang dapat meningkatkan kesuksesan suatu video di platform ini.

Dalam riset yang dilaksanakan oleh Rosita dan Darlin (2022), mengindikasikan jika meningkatkan engagement seperti durasi video, penggunaan hashtag, waktu unggah, dan interaksi dengan audiens sangat memengaruhi tingkat engagement di TikTok. Video yang lebih singkat cenderung menarik perhatian lebih baik, sementara hashtag yang relevan meningkatkan jangkauan [4]. Kondisi ini selaras dengan temuan dalam riset oleh Adam Cahya Dwi Permana et al. (2024), yang mendapati jika konten video dan siaran langsung di TikTok menunjukkan dampak yang menguntungkan pada pilihan konsumen untuk membeli, berdasarkan nilai t-statistik konten video mencapai 6,920 dan signifikansi 0,000, menunjukkan pengaruh yang sangat signifikan terhadap keputusan pembelian [5]. Selain itu, dalam riset yang dilaksanakan oleh Khatimah et al. (2023), ditemukan jika keterlibatan media sosial memberikan dampak positif dan berdampak besar pada loyalitas pelanggan pada e-commerce TikTok Shop. Interaksi aktif di platform sosial, seperti komentar dan berbagi konten, memungkinkan pelanggan membangun hubungan emosional yang kuat dengan merek. Penelitian menunjukkan bahwa konten kreatif dan promosi menarik mendorong partisipasi aktif pengguna, yang pada gilirannya meningkatkan kemungkinan pembelian ulang [6]. Selain itu, temuan dari Odytri Caesar Mahayani et al. (2019) turut mendukung pentingnya pendapatan dalam *social media*. Meskipun fokus penelitian ini adalah pada platform Instagram, prinsip customer engagement yang digunakan tetap relevan untuk platform media sosial lainnya, termasuk TikTok. Dalam studi tersebut, ditemukan bahwa tiga dari empat variabel engagement yakni *self-brand connection*, *company attitude*, dan *word of mouth*—memiliki pengaruh signifikan terhadap kepercayaan pelanggan terhadap merek Shopee. Variabel *company attitude* terbukti menjadi yang paling dominan dalam membangun kepercayaan. Penelitian ini menggunakan metode regresi linear berganda, yang juga relevan untuk mengkaji hubungan antara berbagai faktor dalam engagement video TikTok [7]. Dengan demikian, berdasarkan sejumlah penelitian terdahulu, dapat disimpulkan bahwa berbagai faktor seperti durasi video, waktu unggah, penggunaan hashtag, hingga elemen interaksi aktif (komentar, like, share) berperan penting dalam memengaruhi engagement video di TikTok.

Penelitian ini bertujuan untuk menentukan dan menganalisis berbagai faktor yang mempengaruhi tingkat engagement pada video di platform TikTok. Adapun variabel-variabel yang dianalisis mencakup durasi video, penggunaan hashtag, interaksi dengan pengikut, serta waktu pengunggahan. Penelitian ini juga mengaplikasikan metode Regresi Linear Berganda untuk membangun model prediktif yang mampu memperkirakan tingkat engagement berdasarkan variabel-variabel tersebut. Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi metodologis dalam bentuk model analitik berbasis data, serta kontribusi praktis melalui studi kasus TikTok, sebagai panduan strategis bagi pembuat konten dan pemasar dalam mengoptimalkan performa konten mereka di media sosial.

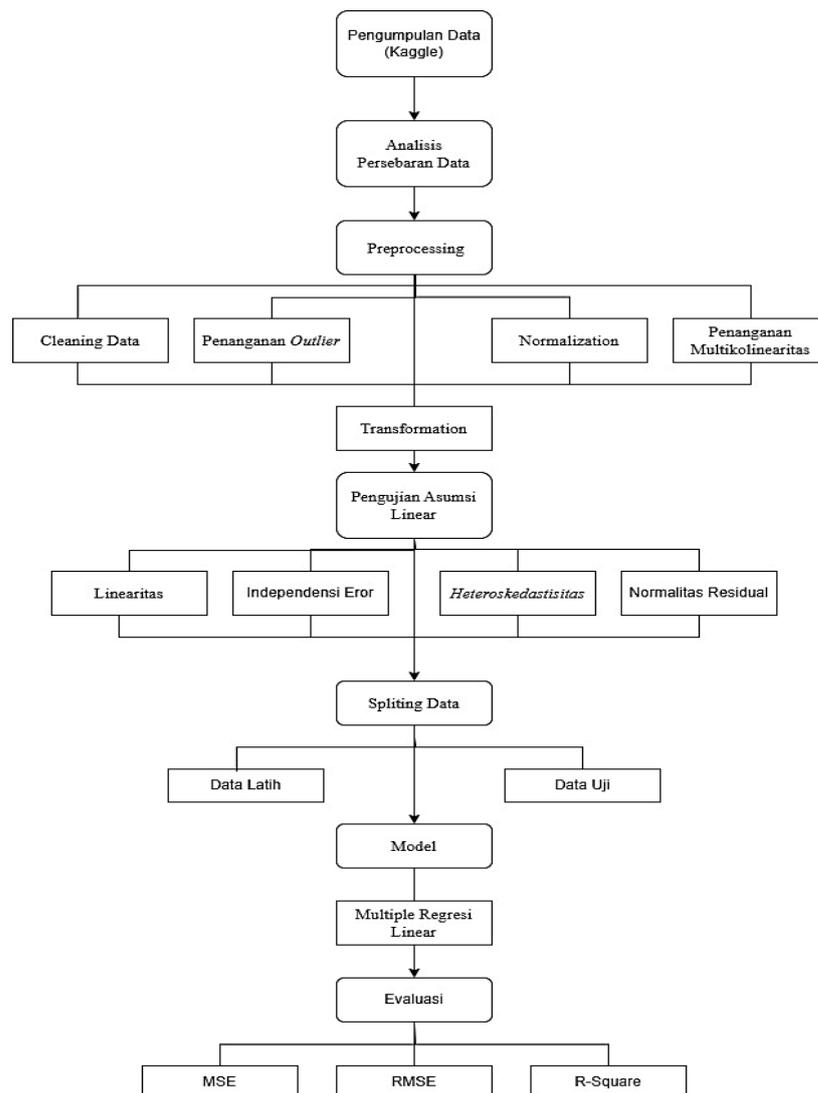
Guna membuat perkiraan sudah pasti perlu sebuah algoritma, satu dia antaranya algoritma untuk membuat perkiraan ini dengan menerapkan *Multiple Regresi Linear*. *Multiple Regresi Linear* adalah suatu prosedur sistematis yang diterapkan guna mengidentifikasi pola korelasi antara faktor dependen dengan lebih dari satu faktor independen [8]. Dalam penerapannya, *Multiple Regresi Linear* sering diterapkan pada bermacam-macam ranah, antara lainnya ekonomi, teknik, dan kajian sosial, untuk memprediksi nilai berdasarkan faktor-faktor yang mempengaruhinya. *Multiple Regresi Linear* sangat sesuai digunakan dalam konteks penelitian ini, mengingat adanya berbagai faktor yang diduga bersama-sama memengaruhi tingkat *engagement*. Selain itu, sebagian besar variabel dalam penelitian ini bersifat kuantitatif dan kontinu, seperti jumlah hashtag, durasi video, dan volume interaksi, yang secara metodologis cocok untuk dianalisis dengan regresi linear. Model ini juga bersifat interpretatif dan prediktif, sehingga mampu memberikan gambaran yang jelas mengenai kontribusi masing-masing variabel terhadap hasil *engagement*.

Hasil akhir dari riset yang sedang dilakukan mengemukakan insight yang lebih dalam terhadap aspek-aspek yang mempengaruhi *engagement* video di *TikTok*, seperti kualitas konten, jumlah pengikut, waktu

posting, pemilihan lagu, dan ajakan untuk bertindak. Selain itu, penelitian ini juga menawarkan panduan bagi pembuat konten dan pemasar untuk memanfaatkan variabel-variabel yang terbukti signifikan dalam meningkatkan *engagement*, sehingga mereka dapat merancang strategi yang lebih efektif dalam menarik perhatian *audiens*. Selanjutnya, riset ini menciptakan kerangka peramalan berbasis pendekatan linier yaitu sesuatu yang dapat diterapkan untuk mengestimasi potensi *engagement* suatu video berdasarkan faktor-faktor yang ada. Maka dari itu, riset yang sedang dilakukan diusahakan mampu menyajikan yang berarti di dalam perencanaan pembuatan konten di *TikTok*, baik untuk individu maupun bisnis yang ingin meningkatkan interaksi dengan *audiens* mereka.

2. METODOLOGI PENELITIAN

Riset ini diterapkan guna merancang pendekatan perkiraan *engagement* video di platform *TikTok* dengan menggunakan metode *Multiple Regresi Linear* Terdapat metode penelitian yang diawali dengan: Pengumpulan Data(1); *Preprocessing*(2); *Splitting Data*(3); Algoritma(4); Model(5). Tahapan Penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 Menjabarkan langkah-langkah riset yang mencakup beberapa tahapan yaitu:

2.1. Pengumpulan Data

Informasi utama untuk riset ini berasal dari set data public seputar *engagement* video pada platform *TikTok* yang disediakan oleh *kaggle*. kumpulan data ini meliputi 19.000 data dengan 9 aspek yang menggambarkan berbagai faktor yang memengaruhi *engagement* video *tiktok* [9].

2.2. Analisis Persebaran Data

Analisis Persebaran Data merupakan tahap awal yang krusial dalam penelitian kuantitatif, di mana peneliti secara sistematis mengeksplorasi karakteristik distribusi setiap variabel dalam dataset. Melalui penggunaan statistik deskriptif (seperti mean, median, standar deviasi) dan teknik visualisasi (seperti histogram dan *box plot*), analisis ini bertujuan untuk mengidentifikasi pola sentralitas, variabilitas, bentuk distribusi, serta keberadaan nilai-nilai ekstrem (*outlier*) [10].

2.3. Preprocessing

Preprocessing data adalah teknik yang digunakan untuk mempersiapkan data demi kemudahan dalam pengolahan atau pemanfaatan. Sasaran tahap prapemrosesan adalah meningkatkan kualitas informasi, meliputi keseluruhan, kestabilan, dan keakuratan waktu, sehingga dapat menghasilkan analisis yang lebih akurat. Proses ini mencakup berbagai langkah, seperti pembersihan data untuk menghilangkan nilai yang hilang atau tidak valid [9].

1. Pembersihan Data

Pembersihan merupakan tahapan dalam mempersiapkan data dengan cara menghilangkan atau melengkapi nilai yang hilang dalam Kumpulan data. Satu di antaranya teknik yang sering diterapkan ialah menggantikan nilai yang kosong dengan rata-rata dari tiap kolom yang bersangkutan [11].

2. Penanganan *Outlier*

Outlier merupakan data yang menyimpang secara mencolok dari pola umum dan dapat memengaruhi hasil analisis secara menyeluruh. *Outlier* dapat disebabkan oleh kesalahan dalam pengukuran atau oleh adanya perbedaan ekstrem dalam populasi data. Untuk menangani *outlier* tersebut, pendekatan statistik seperti metode interkuartil atau *z-score* dapat digunakan untuk mengidentifikasi serta menghilangkan atau menyesuaikan data yang menyimpang [12].

3. *Data Transformation*

Transformasi data merupakan proses penyesuaian informasi agar memperoleh kualitas yang lebih optimal. Proses ini mencakup berbagai teknik, seperti menghilangkan *noise* dari data (*smoothing*), mengagregasi data, melakukan generalisasi, normalisasi, serta membentuk atribut atau fitur baru [13].

4. *Normalization Data*

Normalization Data data adalah salah satu langkah dalam tahap pra-pemrosesan data. Proses ini dilakukan dengan menskalakan kembali nilai-nilai data agar lebih seragam, sehingga dapat meningkatkan efisiensi dan akurasi dalam pemrosesan serta analisis data [14].

5. Penanganan *multikolinearitas*

Multikolinearitas merupakan kondisi ketika variabel bebas saling berkorelasi dan tidak sepenuhnya independen satu sama lain. Indikator yang digunakan untuk mengenali gejala ini adalah *Variance Inflation Factor* (VIF). VIF berfungsi sebagai ukuran yang digunakan dalam mendeteksi multikolinearitas pada model regresi linier yang melibatkan lebih dari dua variabel independen [15].

2.4. Pengujian Asumsi Regresi

Sebelum membangun model regresi linear berganda, asumsi-asumsi klasik perlu diuji pada data yang telah di-*preprocess* untuk memastikan keandalan model.

1. Linearitas

Linearitas merupakan sebuah bentuk pengujian yang memiliki tujuan untuk mengetahui apakah hubungan regresi bersifat linear atau tidak. Pengujian linieritas diterapkan guna mengamati apakah terdapat ketertarikan yang bersifat linier antara variabel independen dan variabel dependen [16].

2. *Independensi Error (Autokorelasi)*

Independensi Error atau autokorelasi diterapkan guna menilai jika dalam konteks model regresi linier ada keterkaitan dalam kesalahan sisa. Bila keterkaitan itu ada, maka kondisi tersebut disebut sebagai masalah *Autokorelasi*. Masalah ini timbul karena observasi yang terjadi secara berkelanjutan seiring waktu [17].

3. *Heteroskedastisitas*

Uji *Heteroskedastisitas* dimaksudkan untuk menilai apabila pada model regresi terdapat ketidakconsistenan varians pada error atau satu observasi dengan observasi lainnya. Bila varians dan

residu bersifat tetap antara pengamatan, maka kondisi ini disebut Homoskedastisitas. Sebaliknya, jika varians tersebut berbeda, maka dinamakan *Heteroskedastisitas* [18].

4. Normalitas Residual

Uji normalitas dilakukan untuk menilai apabila pada model regresi, galat atau residu mengikuti distribusi normal. Salah satu metode statistik non parameterik yang dapat digunakan untuk memeriksa kenormalan residual adalah uji statistik non-parametrik Kolmogorov-Smirnov (K-S) [18].

2.5. *Splitting Data*

Pemisahan data atau pemecahan data merupakan proses signifikan dalam riset yang sedang dilakukan. Tahapan ini memisahkan data ke dalam 2 kelompok, yakni data pelatihan serta data pengujian guna menjamin bahwa model yang dikembangkan mampu bekerja secara optimal pada informasi yang belum pernah diamati sebelumnya. Data pelatihan dimanfaatkan untuk mengajarkan model supaya mampu mengidentifikasi pola dalam informasi tersebut dan memahami hubungan antara variabel independen dan dependen [19]. Data uji membantu dalam menentukan seberapa baik model dapat memprediksi dengan benar [20]. Pada riset ini, pemisahan data diterapkan dengan perbandingan 80% dataset training dan 20% dataset uji coba, yang merupakan pendekatan umum dalam pembelajaran mesin guna memastikan keseimbangan antara pelatihan model dan evaluasi kinerjanya.

2.6. Model

Model dalam pembelajaran mesin merupakan hasil dari proses pelatihan algoritma yang bertujuan untuk mengenali pola, struktur, dan hubungan dalam data. Proses ini dilakukan menggunakan data pelatihan, di mana algoritma bekerja sebagai representasi matematis atau komputasional yang mampu menghasilkan prediksi atau keputusan berdasarkan input baru [21]. Untuk membangun model menggunakan Algoritma *Multiple Regresi Linear*. *Multiple Regresi Linear* memungkinkan analisis keterkaitan antara berbagai faktor dengan engagement video TikTok [22]. Analisis *Multiple Regresi Linier* menggambarkan keterkaitan secara linear antara dua atau lebih variabel bebas (X_1, X_2, \dots, X_p) dengan variabel terikat (Y) [23]. Secara statistik, bentuk umum dari model regresi linier berganda dapat dinyatakan dengan persamaan 1.

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

dimana Y merupakan variabel dependen atau variabel tak bebas yang nilainya diprediksi berdasarkan sejumlah variabel independen atau variabel bebas yang dilambangkan sebagai X_1, X_2, \dots, X_n . Dalam model ini, a adalah konstanta yang menunjukkan nilai Y ketika seluruh variabel independen bernilai nol, sedangkan b_1, b_2, \dots, b_n adalah koefisien regresi yang merepresentasikan besarnya pengaruh masing-masing variabel independen terhadap variabel dependen [24].

2.7. Evaluasi

Tahapan penilaian adalah proses yang bertujuan untuk menyajikan tinjauan terhadap hasil engujian yang telah diterapkan pada sistem. Tahapan penilaian ini memiliki peran penting dalam memahami suatu sistem analisis sentimen serta algoritma yang diterapkan dalam program [25].

1. *Mean Square Error (MSE)*

Kesalahan kuadrat rata-rata merupakan *mean* dari selisih pangkat dua dari selisih nilai yang diprediksi dengan nilai aktual. Meskipun metode ini dapat menghasilkan perbedaan yang cukup besar dalam beberapa kasus, *MSE* umumnya lebih efektif dalam mengukur kesalahan yang kecil, sehingga sering digunakan untuk menilai akurasi model prediksi [26]. Persamaan 2 dapat digunakan untuk menentukan nilai *MSE*.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

dimana n adalah jumlah total data atau observasi, y_i adalah nilai aktual dari data ke- i , dan \hat{y}_i adalah nilai hasil prediksi dari model untuk data ke- i . Simbol \sum menunjukkan proses penjumlahan atas seluruh selisih kuadrat antara nilai aktual dan nilai prediksi.

2. *Root Mean Square Error (RMSE)*

Akar dari Rata-rata Kesalahan Kuadrat merupakan akar dari rata-rata selisih kuadrat antara nilai aktual dan nilai yang diprediksi. Nilai *RMSE* yang lebih kecil mengindikasikan Tingkat kekeliruan yang lebih minim, sehingga model prediksi yang dihasilkan menjadi lebih akurat dan dapat diandalkan [27].

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

dimana n merupakan jumlah data atau observasi, y_i adalah nilai aktual dari data ke- i , dan \hat{y}_i adalah nilai prediksi dari model untuk data ke- i . Simbol \sum menunjukkan penjumlahan dari seluruh selisih kuadrat antara nilai aktual dan nilai prediksi.

3. R-Square

R kuadrat adalah hasil perhitungan pangkat dua dari koefisien korelasi (R). Nilai R kuadrat berada pada rentang nol hingga satu, yang berarti apabila nilai tersebut kecil (mendekati 0), maka hubungan antara kedua variabel cenderung lemah. Sebaliknya, apabila R kuadrat semakin besar (mendekati 1), hubungan antara kedua variabel menjadi semakin kuat [28].

$$R = 1 - \frac{SSE}{SST} \quad (4)$$

dimana SSE (*Sum of Squared Errors*) adalah jumlah kuadrat selisih antara nilai sebenarnya dan nilai prediksi. SST (*Total Sum of Squares*) adalah jumlah kuadrat selisih antara nilai sebenarnya dan rata-rata nilai sebenarnya.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Set data *Kaggle* terdiri dari 19.382 entri video dengan 12 fitur. Dalam penelitian ini, Variabel independen penelitian ini mencakup berbagai faktor yang dapat memengaruhi tingkat *engagement* video di *TikTok*, seperti durasi video, status verifikasi akun, status larangan akun, serta jumlah tayangan, komentar, dan unduhan. Data ini dianalisis menggunakan algoritma regresi linear berganda untuk mengetahui sejauh mana elemen-elemen tersebut berkontribusi terhadap *engagement* video di platform *TikTok*. Sebagai sampel data ditunjukkan pada Table 1.

Table 1. Data Tiktok

	Video duration sec	Claim Status	video share count	Video download count	Video comment count
1	59	Claim	241.0	1.0	0
2	32	Claim	19034.0	11.610	684.0
3	31	Claim	2858.0	833.0	329.0
.....
19.084	58	Claim	281.0	11.0	1.0

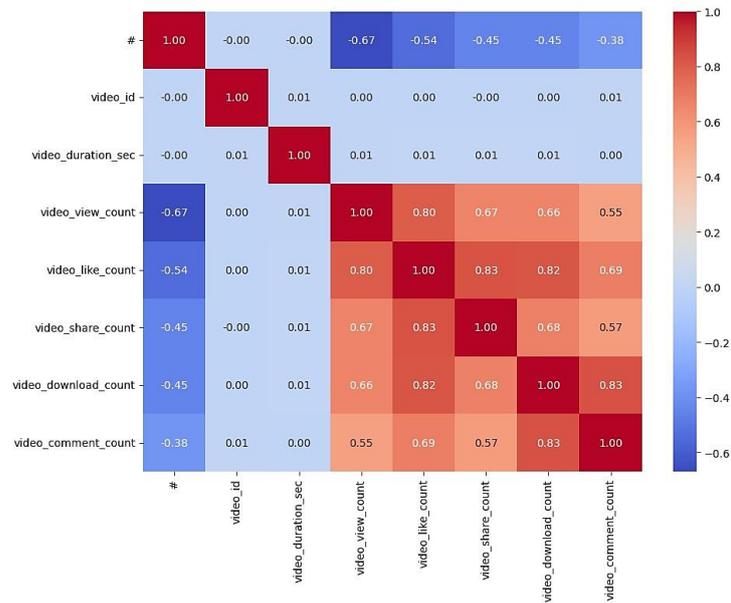
Data tentang *engagement* video di *TikTok* ditunjukkan dalam tabel berikut. Dataset ini mencakup berbagai fitur yang dapat memengaruhi tingkat keterlibatan pengguna terhadap sebuah video. Berikut ini merupakan uraian ringkas dari masing-masing fitur:

1. #: Nomor urut atau indeks dari setiap data dalam dataset.
2. *claim_status*: Status klaim terkait video, menunjukkan apakah video diklaim atau tidak.
3. *video_id*: Identifikasi unik untuk setiap video dalam dataset.
4. *video_duration_sec*: Durasi video dalam satuan detik, menunjukkan berapa lama video tersebut berlangsung.
5. *video_transcription_text*: Teks hasil transkripsi dari video, yang dapat digunakan untuk analisis konten.
6. *verified_status*: Status verifikasi akun pembuat video, apakah akun tersebut sudah terverifikasi atau belum.
7. *author_ban_status*: Status akun pembuat video, apakah terkena larangan atau masih aktif.
8. *video_view_count*: Jumlah total penayangan video, menunjukkan seberapa banyak video telah ditonton.
9. *video_like_count*: Jumlah total suka (*like*) yang diterima video dari pengguna TikTok.
10. *video_share_count*: Jumlah total kali video telah dibagikan oleh pengguna TikTok.
11. *video_download_count*: Jumlah total unduhan video, yang menunjukkan seberapa banyak pengguna menyimpan video tersebut.
12. *video_comment_count*: Jumlah total komentar yang diterima video dari pengguna TikTok

3.2. Analisis Pesebaran Data

Hasil menunjukkan bahwa terdapat korelasi yang sangat kuat antara *video_like_count* dengan *video_share_count* ($r = 0.83$), *video_download_count* ($r = 0.82$), dan *video_view_count* ($r = 0.80$). Selain itu, *video_comment_count* juga menunjukkan hubungan yang kuat dengan *video_download_count* ($r = 0.83$).

Sebaliknya, video_duration_sec memiliki korelasi sangat rendah terhadap semua variabel, mengindikasikan pengaruh yang kecil terhadap interaksi pengguna. Hasil ini menunjukkan adanya hubungan linier positif yang kuat antar beberapa fitur interaksi pengguna, yang berpotensi signifikan dalam model prediksi. Hasil Visualisasi Analisis Pesebaran Data dapat dilihat pada Gambar 1.



Gambar 1. Visualisasi Analisis Pesebaran Data

3.3. Preprocessing

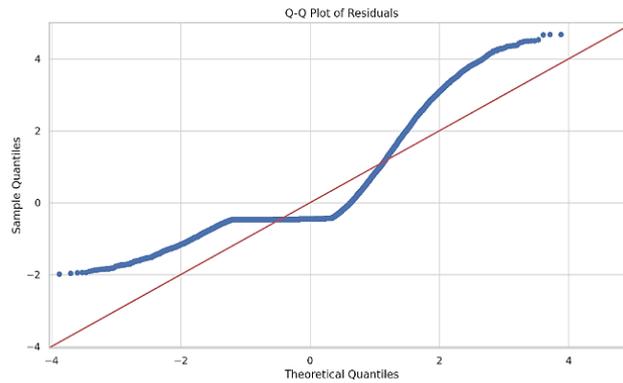
Hasil dari preprocessing, jumlah data berkurang dari 19.382 menjadi 19.084 baris akibat penghapusan 298 data kosong. Data kategorikal seperti claim_status, verified_status, dan author_ban_status telah diubah menjadi nilai numerik (0 dan 1), serta video_transcription_text dikodekan menggunakan indeks angka. Normalisasi dengan metode Min-Max dilakukan pada kolom claim_status, menghasilkan nilai 1.0 pada seluruh data. Deteksi dan penghapusan outlier menggunakan metode IQR dilakukan pada kolom numerik, dengan total 8.675 data dibuang: 1.726 dari video_like_count, 2.447 dari video_share_count, 2.596 dari video_download_count, dan 1.906 dari video_comment_count. Hasil pengujian multikolinearitas menunjukkan bahwa semua nilai VIF berada di bawah 5, dengan nilai tertinggi 1.73 pada video_download_count, sehingga tidak ditemukan indikasi multikolinearitas antar variabel. Hasil preprocessing data dapat dilihat pada tabel 2.

Table 2. Hasi Preprocessing Data

claim_status	video_id	video_duration_sec	video_transcription_text	verified_status	author_ban_status	video_view_count	video_like_count	video_share_count	video_download_count	video_comment_count	claim_status_normalized
1	7,018E+09	59	0	0	0.0	343296.0	19425.0	241.0	1.0	0.0	1.0
2	4,014E+09	32	1	0	1.0	140877.0	77355.0	19034.0	1161.0	684.0	1.0
3	9,86E+09	31	2	0	1.0	902185.0	97690.0	2858.0	833.0	329.0	1.0
4	1,867E+09	25	3	0	1.0	437506.0	239954.0	34812.0	1234.0	584.0	1.0
5	7,105E+09	19	4	0	1.0	56167.0	34987.0	4110.0	547.0	152.0	1.0

3.4. Pengujian Asumsi Regresi

Pengujian asumsi regresi linear menunjukkan bahwa model memenuhi asumsi linearitas, tidak adanya autokorelasi (Durbin-Watson = 2.007), dan tidak terdapat multikolinearitas (seluruh nilai VIF < 5). Namun, asumsi normalitas residual dan homoskedastisitas belum terpenuhi, sebagaimana ditunjukkan oleh hasil uji Kolmogorov-Smirnov dan Breusch-Pagan dengan nilai $p < 0,05$, meskipun telah dilakukan transformasi logaritmik. Mengingat ukuran data yang besar serta tujuan utama penelitian ini adalah prediksi, bukan inferensi statistik, maka pelanggaran terhadap kedua asumsi tersebut masih dapat ditoleransi. Sebagai langkah lanjutan, disarankan penggunaan regresi dengan robust standard errors atau pendekatan non-parametrik untuk meningkatkan keandalan model yang dapat dilihat pada Gambar 2.



Gambar 2. Pengujian Asumsi Regresi

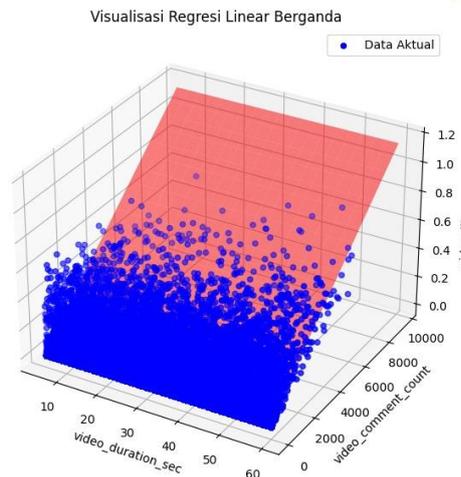
3.5. Splitting Data

Riset ini membagi Kumpulan data ke dalam tiga tahap dengan perbandingan 80:20, setelah data dinormalisasi menggunakan *MinMaxScaler*. Pada tahap ini, 80% Sebagian dari keseluruhan data dimanfaatkan sebagai data pelatihan, sementara 20% dimanfaatkan sebagai data pengujian. Pembagian data hal demikian memungkinkan analisis performa model dalam berbagai skenario di mana proporsi data latih lebih besar dibandingkan data uji, sehingga model dapat belajar lebih banyak sebelum diuji. Data yang telah diproses ini kemudian dapat digunakan untuk pelatihan model prediksi Regresi Linier Linear. untuk menganalisis faktor-faktor yang mempengaruhi jumlah like pada video *TikTok*.

3.6. Pemodelan Regresi Linear Berganda

Model regresi linear berhasil diterapkan untuk memprediksi jumlah tampilan video (*video_view_count*). Dari hasil Feature Selection, ditemukan bahwa fitur *video_like_count*, *video_share_count*, dan *video_download_count* memiliki pengaruh paling besar terhadap jumlah tampilan. Model ini memungkinkan peneliti untuk mengidentifikasi dan mengukur kontribusi masing-masing variabel independen terhadap variabel dependen (*engagement*), sehingga hasilnya dapat ditafsirkan secara langsung.

Pada grafik 3D Gambar 3, bidang merah mewakili model regresi linear berganda yang dipelajari dari data yang diberikan. Sumbu X menunjukkan *Jam Menonton* (atau faktor lain seperti *Durasi Video* jika datanya berbeda), sumbu Y menunjukkan *Jumlah Komentar*, dan sumbu Z menunjukkan *Jumlah Likes*.



Gambar 3. Hasil Pemodelan Regresi Linear Berganda

Dari bidang regresi, terlihat bahwa jumlah komentar (X_2) memiliki pengaruh besar terhadap jumlah likes. Ini berarti semakin banyak komentar yang diterima suatu video, semakin tinggi jumlah likes yang diprediksi oleh model. Pola data menunjukkan bahwa hubungan antara jumlah komentar, jam menonton, dan jumlah likes dapat direpresentasikan dengan model linier. Variabel *video_comment_count* memiliki koefisien sebesar 0.85 ($p < 0.001$), menunjukkan pengaruh signifikan terhadap *video_like_count*. Sebaliknya, *video_duration_sec* memiliki koefisien 0.02 ($p = 0.22$), yang tidak signifikan pada tingkat kepercayaan 95%.

3.7. Evaluasi

3.7.1. Mean Square Error (MSE)

MSE sebesar 0.005502 menunjukkan bahwa rata-rata selisih kuadrat antara nilai prediksi dan nilai sebenarnya cukup kecil. Hasil ini mengindikasikan bahwa model regresi linear yang digunakan memiliki tingkat kesalahan yang rendah dalam memprediksi jumlah tampilan video, sehingga dapat dikatakan cukup akurat dan dapat diandalkan untuk estimasi jumlah tampilan berdasarkan fitur yang dipilih. Meskipun demikian, masih ada peluang untuk meningkatkan akurasi model dengan menambahkan fitur lain yang mungkin berpengaruh atau menggunakan model yang lebih kompleks jika diperlukan.

3.7.2. Root Mean Square Error (RMSE)

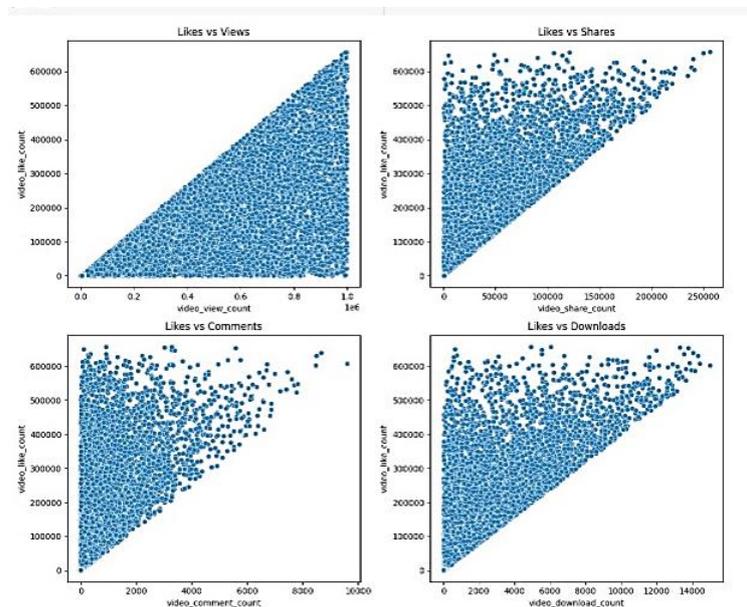
Root Mean Squared Error (RMSE) sebesar 0.0742 menunjukkan bahwa rata-rata kesalahan prediksi model regresi linear terhadap nilai sebenarnya adalah sekitar 7.42% dalam skala data asli. Nilai RMSE yang kecil menandakan bahwa model memiliki tingkat error yang rendah, sehingga dapat dikatakan cukup akurat dalam memprediksi jumlah tampilan video. Keakuratan model ini menunjukkan bahwa regresi linear mampu menangkap pola hubungan antara fitur yang digunakan dengan jumlah tampilan video secara cukup baik.

3.7.3. R-square (R^2)

Nilai R-squared (R^2) sebesar 0.978 menunjukkan bahwa 97,8% variasi jumlah like pada video TikTok dapat dijelaskan oleh variabel-variabel independen* yang digunakan dalam model. Ini menandakan bahwa model regresi linear yang dibangun memiliki akurasi prediksi yang sangat tinggi dan mampu merepresentasikan hubungan antara fitur-fitur seperti durasi, komentar, share, download, dan view terhadap jumlah like secara efektif.

3.8. Analisis Faktor

Analisis faktor yang mempengaruhi *engagement* video tiktok jumlah *views* memiliki pengaruh paling kuat terhadap jumlah likes, diikuti oleh *shares*, *comments*, dan *downloads*. Semakin tinggi jumlah *views*, semakin tinggi pula jumlah likes, dengan korelasi yang kemungkinan mendekati 0.9. Sementara itu, *shares* dan *comments* juga berkontribusi dengan korelasi sekitar 0.6 - 0.7, sedangkan *downloads* memiliki pengaruh yang lebih kecil sekitar 0.5 - 0.6. Dengan demikian, untuk meningkatkan jumlah likes, faktor utama yang harus ditingkatkan adalah jumlah *views*, diikuti oleh interaksi melalui *shares* dan *comments*. Visualisasi dari analisis faktor dapat dilihat pada gambar 4.



Gambar 4. Hasil Visualisasi

3.9. Diskusi Analisis

Hasil penelitian menunjukkan bahwa jumlah tayangan, like, komentar, share, dan durasi video berpengaruh signifikan terhadap tingkat *engagement* video TikTok. Temuan ini sejalan dengan penelitian sebelumnya oleh Nur Sapina et al. (2024) dan Permana et al. (2024), yang menemukan bahwa interaksi pengguna seperti komentar dan share sangat memengaruhi performa konten di media sosial. Khatimah et al. (2023) juga menunjukkan bahwa *engagement* memiliki hubungan dengan loyalitas pelanggan dalam konteks pemasaran digital.

Namun, penelitian ini masih memiliki beberapa keterbatasan. Belum dilakukan analisis persebaran data seperti outlier atau distribusi normal pada tiap variabel. Selain itu, belum dilakukan pengujian multikolinieritas antar variabel independen, yang bisa memengaruhi keakuratan model regresi. Asumsi-asumsi dasar regresi seperti normalitas residual, heteroskedastisitas, dan autokorelasi juga belum diuji. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh content creator dan pemasar untuk menyusun strategi konten yang lebih efektif, misalnya dengan meningkatkan interaksi melalui komentar atau share. Model regresi ini juga bisa digunakan untuk memperkirakan performa konten sebelum dipublikasikan.

Untuk penelitian selanjutnya, disarankan menambahkan analisis lanjutan terhadap distribusi data, pengujian asumsi regresi, serta mempertimbangkan variabel tambahan seperti waktu unggah, jenis konten, atau sentimen. Metode lain seperti Random Forest atau XGBoost juga bisa dipertimbangkan agar hasil prediksi lebih akurat.

4. KESIMPULAN

Riset yang dilakukan terhadap Kumpulan data *Tiktok* yang diambil dari *Kaggle*, yang terdiri dari 19.000 data, berhasil mengkaji unsur-unsur yang berpengaruh tingkat engagement video di platform *TikTok*. Dengan menggunakan metode *Multiple Regresi Linear*, penelitian ini menemukan bahwa faktor-faktor seperti durasi video, jumlah tayangan, jumlah *like*, jumlah *share*, dan jumlah komentar memiliki pengaruh yang signifikan terhadap tingkat *engagement* video. Hasil analisis menunjukkan bahwa jumlah tayangan (*video view count*) memiliki pengaruh yang paling besar terhadap jumlah *like*, *share*, dan komentar, yang pada gilirannya memengaruhi keseluruhan tingkat *engagement*. Model prediksi yang dibangun menggunakan algoritma regresi linear menunjukkan bahwa jumlah *like*, *share*, dan *download* video juga memiliki kontribusi yang penting dalam memprediksi tingkat *engagement*. Berdasarkan evaluasi model menggunakan *MSE* dan *RMSE*, ditemukan bahwa model ini cukup akurat dalam memprediksi jumlah tampilan dan tingkat *engagement* video dengan tingkat kesalahan yang relatif rendah.

Secara keseluruhan, temuan dari penelitian ini memberikan wawasan yang berharga bagi pembuat konten dan pemasar untuk merancang strategi yang lebih efektif dalam meningkatkan *engagement* video di *TikTok*. Penggunaan model prediksi berbasis regresi linear ini diharapkan dapat membantu individu dan bisnis dalam mengoptimalkan konten mereka untuk mencapai interaksi yang lebih tinggi dengan *audiens*. Penelitian ini juga membuka peluang untuk penelitian lebih lanjut dengan menambahkan lebih banyak variabel atau menggunakan metode analisis yang lebih kompleks guna meningkatkan akurasi prediksi.

REFERENSI

- [1] A. Cahya, D. Permana, P. Anugrah, and C. Dewi, "Analisis Pengaruh Konten Video Dan Siaran Langsung Pada Tiktok Terhadap Keputusan Pembelian Menggunakan Metode Regresi Linear Berganda (Studi Kasus Toko Busana HI Outfit)," vol. 4, pp. 8961–8977, 2024.
- [2] R. Rosita and Evalina Darlin, "Pengaruh Kualitas Konten Tik Tok Terhadap Customer Engagement Pada Customer Queensha," *Jurnal Lentera Bisnis*, vol. 13, no. 2, pp. 1061–1071, 2024, doi: 10.34127/jrlab.v13i2.1129.
- [3] Subadi, "Analisa Engagement Media Sosial Terhadap Viralnya Film Pendek Tilik (2018)," Mar. 2022.
- [4] R. Rosita and Evalina Darlin, "Pengaruh Kualitas Konten Tik Tok Terhadap Customer Engagement Pada Customer Queensha," *Jurnal Lentera Bisnis*, vol. 13, no. 2, pp. 1061–1071, May 2024, doi: 10.34127/jrlab.v13i2.1129.
- [5] A. C. D. Permana, H. S. Alam, and P. A. C. Dewi, "Analisis Pengaruh Konten Video Dan Siaran Langsung Pada Tiktok Terhadap Keputusan Pembelian Menggunakan Metode Regresi Linear Berganda (Studi Kasus Toko Busana HI Outfit)," *Innovative: Journal Of Social Science Research*, vol. 4, no. 4, pp. 8961–8977, 2024.
- [6] H. Khatimah, M. R. Basalamah, and T. S. Ramadhan, "Pengaruh Strategi Content Marketing, Engagement Media Sosial Dan Customer Experience Terhadap Loyalitas Pelanggan Pada E-Commerce Tiktok Shop (Studi Kasus Pada Mahasiswa FEB Unisma)," *e – Jurnal Riset Manajemen*, vol. 14, 2025.
- [7] O. C. Mahayani, I. Aknuranda, and A. Kusyanti, "Pengaruh customer engagement melalui media sosial terhadap kepercayaan merek (Studi Kasus: Instagram Shopee)," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 3, no. 4, pp. 3301–3310, 2019.
- [8] D. K. S. Nadeak, F. Taufik, and M. Iswan, "Estimasi Produksi Dolomit Untuk Pakan Ternak Udang Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 4, pp. 527–536, 2023.
- [9] J. Homepage, I. Permana, and F. Nur Salisah, "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation Terhadap".
- [10] A. Sholikhah, "Statistik deskriptif dalam penelitian kualitatif," *KOMUNIKA: Jurnal Dakwah Dan Komunikasi*, vol. 10, no. 2, pp. 342–362, 2016.

- [11] D. Nabila Herisnan, E. Dadyrata, and L. Efrizoni, "Komparasi Algoritma Decision Tree, Svm, Naive Bayes Dalam Prediksi Penyakit Liver," 2024. [Online]. Available: <http://ojsamik.amikmitragama.ac.id>
- [12] N. Nurussakinah and M. Faisal, "Klasifikasi penyakit diabetes menggunakan algoritma decision tree," *Jurnal Informatika*, vol. 10, no. 2, pp. 143–149, 2023.
- [13] Gde Agung Brahmama Suryanegara, Adiwijaya, and Mahendra Dwifabri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 114–122, Feb. 2021, doi: 10.29207/resti.v5i1.2880.
- [14] B. Prasajo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 79–89, Sep. 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.
- [15] M. Sriningsih, D. Hatidja, and J. D. Prang, "Penanganan multikolinearitas dengan menggunakan analisis regresi komponen utama pada kasus impor beras di Provinsi Sulut," *Jurnal Ilmiah Sains*, pp. 18–24, 2018.
- [16] S. Syahidin and E. Erma, "Analisis Faktor-Faktor Yang Mempengaruhi Produksi Sere Wangi Di Kecamatan Terangun Kabupaten Gayo Lues," *Gajah Putih Journal of Economics Review*, vol. 3, no. 2, pp. 76–88, 2021.
- [17] S. Prasetyo, "Pengaruh Red Flags, Skeptisme Profesional Auditor, Kompetensi, Independensi, Dan Profesionalisme Terhadap Kemampuan Auditor Dalam Mendeteksi Kecurangan (Studi Empiris Pada Kantor Akuntan Publik Di Pekanbaru, Padang, Dan Medan Yang Terdaftar Di Iapi 2013)," 2015.
- [18] C. Wulandari and D. Efendi, "Pengaruh profitabilitas terhadap nilai perusahaan dengan corporate social responsibility sebagai variabel moderasi," *Jurnal Ilmu Dan Riset Akuntansi (JIRA)*, vol. 11, no. 6, 2022.
- [19] U. Prediksi, P. Bantuan, J. Kesehatan, N. Attamami, A. Triayudi, and R. T. Aldisa, "Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 Teknologi Komunikasi dan Indormatika," vol. 7, no. 2, 2023, doi: 10.35870/jti.
- [20] F. Syahro and N. Fitriani, "Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, Dan K-Nearest Neighbors Dalam Prediksi Harga Saham," *Jar's*, vol. 2, no. 1, p. 13, [Online]. Available: <https://www.ejournalwiraraja.com/index.php/JARS>
- [21] A. D. Putri, F. Sholekhah, E. Dadyrata, L. Efrizoni, R. Rahmaddeni, and N. Sapina, "Penerapan Algoritma Decesion Tree C4.5 untuk Memprediksi Tingkat Kelangsungan Hidup Pasien Kanker Tiroid," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1485–1495, Sep. 2024, doi: 10.57152/malcom.v4i4.1532.
- [22] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *Jurnal Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, May 2023, doi: 10.24246/juses.v6i1p34-43.
- [23] N. Afrilia and F. Frazna Az-Zahra, "Prediksi Hasil Panen Wortel Menggunakan Algoritma Regresi Linear Berganda," 2024.
- [24] D. M. Huda, G. Dwilestari, A. R. Rinaldi, and I. Solihin, "Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 150–157, 2024.
- [25] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [26] M. Fikri, I. Raehan, A. Budiman Kusdinar, and D. Indrayana, "Penerapan Regresi Linier Berganda Untuk Memprediksi Hasil Panen Kacang Kedelai (Studi Kasus: Kecamatan Surade)," 2024.
- [27] J. Pebralia, "JIFP (Jurnal Ilmu Fisika dan Pembelajarannya) Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook Rainfall Analysis using Machine Learning-Multiple Linear Regression Method Based on Python and Jupyter Notebook," vol. 6, no. 2, pp. 23–30, 2022, [Online]. Available: <http://jurnal.radenfatah.ac.id/index.php/jifp/>
- [28] N. Muktiadji and R. Kamage, "Pengaruh Penjualan Dan Profitabilitas Terhadap Pertumbuhan Perusahaan Studi Kasus Pada PT. Gudang Garam, Tbk. Dan PT. Bentoel Internasional Investama, Tbk," *Jurnal Ilmiah Ranggagading*, vol. 9, no. 1, pp. 1–6, 2009.